



# UNIVERSIDAD DE LA RIOJA

## TRABAJO FIN DE ESTUDIOS

Título

Instruments for Assessing Vocabulary in English as a Foreign Language

Autor/es

PAULA ESTEBAN VIANA

Director/es

M.<sup>a</sup> PILAR AGUSTÍN LLACH

Facultad

Facultad de Letras y de la Educación

Titulación

Grado en Estudios Ingleses

Departamento

FILOLOGÍAS MODERNAS

Curso académico

2017-18



***Instruments for Assessing Vocabulary in English as a Foreign Language***, de  
PAULA ESTEBAN VIANA

(publicada por la Universidad de La Rioja) se difunde bajo una Licencia Creative  
Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported.

Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los  
titulares del copyright.

© El autor, 2018

© Universidad de La Rioja, 2018

[publicaciones.unirioja.es](http://publicaciones.unirioja.es)

E-mail: [publicaciones@unirioja.es](mailto:publicaciones@unirioja.es)



# TRABAJO FIN DE GRADO

## Título

**Instruments for Assessing Vocabulary in English as a Foreign Language**

---

## Autor

Paula Esteban Viana

---

## Tutor/es

María del Pilar Agustín Llach

---

## Grado

Grado en Estudios Ingleses [601G]

---

**Facultad de Letras y de la Educación**

Año académico

2017/18

**INSTRUMENTS FOR ASSESSING VOCABULARY IN ENGLISH AS A  
FOREIGN LANGUAGE**

**INSTRUMENTOS PARA EVALUAR EL VOCABULARIO EN INGLÉS COMO  
LENGUA EXTRANJERA**

**ABSTRACT**

This work is an overview of the main instruments available to assess vocabulary in a foreign language. In order to do so, the first section offers a short introduction about the main purpose that will be discussed and how it will be developed. The following point proposes several definitions of basic terms that are relevant to the topic. Then, the third section tries to explain what it means to know a word and in what different ways a word can be known. Next, section four attempts to define lexical competence and provides a summary of one of the tools that exists to measure general competence in a foreign language. The following section explains some of the factors that influence the assessment of vocabulary. Finally, section six tries to clarify the different types of tests and instruments available to evaluate vocabulary, focusing mainly on the difference between the breadth and depth knowledge of vocabulary in a foreign language.

Este trabajo es un repaso de los principales instrumentos disponibles para evaluar el vocabulario en lengua extranjera. Para ello, la primera sección ofrece una pequeña introducción sobre el tema que se va a tratar y cómo se va a desarrollar. El siguiente punto propone varias definiciones de términos básicos relevantes para el tema. Después se intenta explicar que significa conocer una palabra y que formas diferentes de hacerlo existen. En el punto cuatro se define que es la competencia léxica y se hace un resumen de una de las herramientas que existen para medir la competencia en lengua extranjera. En el siguiente apartado se explican algunos de los factores que influyen en la evaluación de vocabulario. Por último, el punto seis trata de esclarecer los diferentes tipos de test e instrumentos que existen para evaluar el vocabulario, centrándose principalmente en la diferencia entre la amplitud y la profundidad con la que se conocen las palabras en el idioma extranjero.

## **TABLE OF CONTENTS**

### **1. Introduction**

### **2. Definition of Basic Terms**

2.1 Some Definitions of Word

2.2 Lexical Item or Lexeme, Lexicology and Lexicography

2.3 Types and Tokens

2.4. Word-Formation Process

### **3. Word Knowledge: What does it Mean to Know a Word?**

### **4. Factors which Affects Vocabulary Measuring**

4.1 Reliability and Validity

4.2 Word Frequency and Frequency Model

4.3 Word Difficulty

### **5. Testing Vocabulary**

5.1 Testing and Assessment

5.2 Testing Vocabulary

### **6. Conclusions**

### **7. References**

### **Appendix 1**

### **Appendix 2**

## 1. Introduction

Nowadays it is very important to have a good command of, at least, one foreign language in order to be competent in a job or to travel to a foreign country and communicate with people without any problem. It seems obvious that one of the preferred foreign or second language to learn has always been English. There are several methodologies to teach languages as well as several ways to assess the level of learners in that language. So, this work tries to identify the different tools or kind of exercises that teachers or researchers in this field can use to measure English competence, focusing mainly in vocabulary knowledge, which is an essential part for mastering a language. Vocabulary assessment has an important role not only in determining the level in which the learners are, but also for developing textbooks, syllabus, course content, etc. based on the vocabulary tests scores.

## 2. Definition of Basic Terms

### 2.1 Some Definitions of Word

The term *word* is quite challenging to define in an accurate way because it may have several meanings. As Trask states (1997: p. 228) “a word is a linguistic unit typically longer than a morpheme but smaller than a phrase”, and it can be defined, at least, in 4 levels: orthographic word, phonological word, lexical item or lexeme and grammatical word-form or morphosyntactic word.

The linguist David Crystal in *A Dictionary of Linguistics and Phonetics* (2008: p. 522) defines the term *word* in three different ways, and he also comments that it has several definitions depending on the point of view.

- (a) Words are the physically definable units which one encounters in a stretch of writing (bounded by spaces) or speech (where identification is more difficult, but where there may be phonological clues to identify boundaries, such as a pause, or juncture features). ‘Word’ in this sense is often referred to as the orthographic word (for writing) or the phonological word (for speech). A neutral term often used to subsume both is word form.
- (b) There is a more abstract sense, referring to the common factor underlying the set of forms which are plainly variants of the same unit, such as walk, walks, walking, walked. The ‘underlying’ word unit is often referred to as a lexeme. Lexemes are the units of vocabulary, and as such would be listed in a dictionary.

(c) This then leaves the need for a comparably abstract unit to be set up to show how words work in the grammar of a language, and ‘word’, without qualification, is usually reserved for this role (alternatively, one may spell out this implication, referring to ‘morphemic/morphosyntactic/grammatical’ words, though the latter has an alternative sense). A word, then, is a grammatical unit, of the same theoretical kind as morpheme and sentence. In a hierarchical model of analysis, sentences (clauses, etc.) consist of words, and words consist of morphemes (Crystal, 2008: p. 522).

According to the definition provided by Bloomfield, a word is “a minimal free form, i.e. the smallest unit which can constitute, by itself, a complete utterance” (Crystal, 2008). This criterion is not widely accepted by linguists and researchers in the field, and there is some disagreement about the accuracy of the definition.

Some linguists have also offered a subclassification of words based on, for instance, the distinction between function and content or lexical words, closed-class and open-class words, empty and full words, etc. Then, words can be specifically grouped in word-classes by analyzing their grammatical, semantic and phonological properties.

## 2.2 Lexical Item or Lexeme, Lexicology and Lexicography

As Crystal (2008: p. 279) writes in his dictionary, “a unit of vocabulary is generally referred to as a *lexical item* or *lexeme*.” A dictionary contains all the lexical items in a language or the lexicon, and these items are listed as a set of lexical entries.

A lexical set includes a group of words which share some semantic or formal features. “A group of items used to identify the network of contrasts in a specific semantic or lexical field (e.g. cooking, color) may also be called a lexical system.”

It is also important to bear in mind the distinction between two branches of linguistics; lexicology and lexicography.

*Lexicography* is a branch of applied linguistics which consists in observing, collecting, selecting, and describing units from the stock of words and word combinations in one or more languages. Lexicography also includes the development and description of the theories and methods which are to be the basis of this activity.

*Lexicology* is regarded as the branch of linguistics which deals with the study of vocabulary, its structure and other characteristics. This refers, first of all, to the meaning

of words and the relationship between meanings (semantics). Besides, lexicology studies the formation and structure of individual words (word-formation or morphology).

### 2.3 Types and Tokens

Sometimes the term *word* is used in its general sense, and people tend to use it when they actually refer to terms such as types, tokens, lemmas, word-families, etc. It is important to have these concepts clearly defined in our mind above all, when we want to make vocabulary assessment tests.

*Tokens* can be defined as the total number of words that a sentence or a text contain whereas *types* are the number of different words that appear in a text. For instance, in the sentence *The cat is on the table* there are six tokens and five types since the word “the” appears twice.

In the vocabulary tests used some years ago in order to measure vocabulary knowledge, words such as *help*, *helped* or *helps* counted as different words but, with the passing of time, and for greater accuracy, test such as that of Goulden *et al.* (1990) treated the common inflections and derivational forms as a single unit, so, in the previous example, we would count it as only one word.

Thus, if we want to know the total number of words a student has written in an essay or in a written exam, the first count (based on tokens) would be useful, but if we want to assess the real vocabulary that she or he has available, we have to discard words which are repeated, inflections and derivations or words included in the same word-family.

When dealing with written samples it is quite straightforward to do this sort of count, but it is more difficult when we focus on oral production. As Milton (2009) states, “what to count as a word may be a difficult task.” Should expressions such as *ums* or *ers* count as words? And how do we count contractions such as *isn't*, *don't* or *won't*? There is not a unique response to these questions, it depends on the type of tests, the research that is going to be conducted or the data we want to collect.

### 2.4 Word-formation

In its general sense, word-formation could be defined as the process by which we create new words from the existing ones. It helps users of a language to broaden their vocabulary knowledge in an economical way. This process includes not only inflectional (*car-cars*)



and derivational forms (*use-useful* or *regard-disregard*), but also more recent processes of constructing new words, some of them are mentioned below.

In *compounding* a novel word is formed by joining two (or more) existing words, as in *blackboard*, *eco-friendly* or *nowadays*. Compounds can be made in all word classes.

According to Bauer (2004: p. 326), “*clipping* is the process of shortening a word without changing its meaning or its part of speech”. Some linguists distinguish between three types of clipping: *back-clipping*, where the beginning of the word, usually the first syllable, is maintained as in *examination-exam* or *demonstration-demo*. The second type is called *fore-clipping*, here the last part of the word is kept as in *telephone-phone* or *airplane-plane*. The third type is not as frequent as the other two mentioned above, it is called *middle clipping* and it consists in eliminate the beginning and the end of a word in order to create the new form as in *refrigerator-fridge* or *influenza-flu*.

The last process to focus on is called *blending*. A new word is created by combining parts of two different words. Some examples are *smog* (from smog + frog) or *sitcom* (situation + comedy).

English native speakers tend to use these words that come from word-formation processes very frequently in their daily life speech, but they are not aware due to the fact that these words are deeply internalized in their brains.

### **3. Word Knowledge: What does it Mean to Know a Word?**

The notion of word knowledge is very closely related with the notion of word that we take into account when making estimations about the vocabulary size of a certain individual. Following Milton (2009), it is very common to classify word knowledge into *receptive* or *passive* and *productive* or *active*. But, what are the differences between these types of knowledge?

*Receptive* knowledge makes reference to the words a learner recognized when listening or reading them and *productive* knowledge are the words that a learner can use in his or her oral or writing production.

In addition, many material and curriculum designers divide the books in words that the student should know receptively and the ones they should only know passively. For instance, according to Krizsán (2003) the Hungarian Core Curriculum suggests that

“learners should learn some 1600 words by the 8<sup>th</sup> grade, of which 1200 should be known actively and a further 400 passively” (Milton, 2009).

Another important distinction is the one suggested by Anderson and Freebody (1981), they distinguish between *breadth* of word knowledge and *depth* of word knowledge. *Breadth* is basically the number of words a person knows and *depth* has to do with the concrete knowledge of each word, but I will explain these notions in detail later in this work.

As it can be observed in the following table, Nation (2001: p. 27) tries to provide a more complete and accurate division of word knowledge. In order to accomplish this task, he separates word knowledge into three distinct areas: knowledge of *form*, knowledge of *meaning* and knowledge of *use*.

Form	Spoken	R
		P
	Written	R
		P
	Word parts	R
		P
Meaning	Form and meaning	R
		P
	Concepts and referents	R
		P
	Associations	R
		P
Use	Grammatical functions	R
		P
	Collocations	R
		P

	Constraints on use	R
		P

Table 1. What is involved in knowing a word (Nation, 2001: p. 27)

As it is shown in this table, word knowledge is divided into form, meaning and use, and then, these parts are subclassified. Within this subclassification, each part has a receptive and a productive area.

Following this, a receptive *spoken* knowledge of a word *form* would be to know what the word sounds like, and a productive spoken knowledge of a word form would involve to know how the word is pronounced, the phonological form. Then, the *written* knowledge of a word form involves knowing what a word looks like and how it is spelled. And finally, regarding the *word parts*, Nation makes reference to the area of knowledge by which a learner knows the affixes that can be add to a word in order to change its meaning.

Knowledge of word *meaning* is likewise divided into three. The first part, *form and meaning* implies the capacity of linking the form and the meaning of a certain word. Then, the other sub-divisions, *concepts and referents* and *associations* point out that a word can have different translations and associations in one language.

For example, a word such as *fat* carries very negative connotations in English when describing a person. But, in parts of southern Nigeria, the concept has historically had very positive connotations. Language learners need to know this kind of information if they are not to cause confusion or offence by the wrong choice of words (Milton, 2009: p. 15).

Finally, knowledge of word *use* is divided into *grammatical functions*, *collocations* and *constraints of use*. The first one implies knowing what part of speech a word is and how it links with other words. Collocations refers to the company words like ‘to keep’, words that frequently collocates next to each other. The last sub-section has to do with the restrictions of words, some words are highly restricted in their company and others occurs very frequently.

To finish this section, it is very important to define the concept of *lexical competence*. Authors have referred to this concept using many different terms such as *knowledge of a word*, *knowledge of vocabulary*, *knowledge of lexicon* and *semantic competence*. A general definition could be the knowledge and use of a word by a speaker of language.

According to Meara (1996), “lexical competence might be described in terms of a very small of easily measurable dimensions. These dimensions are properties of the lexicon considered as a whole.” Some of the basic dimensions are thought to be size of vocabulary (the number of words a learner knows) and organization. (Meara, 1996: p. 10)

Traditionally, lexical competence has been defined in terms of the “specification of all the knowledge that speakers might have about words in their lexicon” (Meara, 1996: p. 14). This is a very plausible definition in theory but difficult to develop in practice because “you can only develop models of lexical competence of this sort if you have a complete model of semantics and a complete specification of the syntactic and associational behavior of all the words in a speaker’s lexicon” (Meara, 1996: p.14).

Jiménez Catalán (2002: p. 8), divides the different studies that have provided a definition, a description and a list of the dimensions of lexical competence in two groups, the linguistic perspective, which analyzes the complexity of meaning and structure, and the psycholinguistic perspective, reflected in the analysis of the difficulties in acquiring lexical competence and the other competences and skills involve in learning a language.

#### **4. Factors which Affects Vocabulary Measuring**

It is obvious that the importance of vocabulary knowledge in the acquisition of a foreign language has increased in the past decades. Thus, it is necessary to find an accurate and valid way to test the learner’s vocabulary acquisition. According to Lewis (2008), “the single most important task facing language learners is acquiring a sufficiently large vocabulary”.

##### **4.1 Reliability and Validity**

Until the last twenty years or so, vocabulary tests were not standardized and they “had to be created *at hoc*” (Milton, 2009: p. 17). Nowadays, researchers and teachers of foreign languages try to create a common set of vocabulary tests that condense all the factors that mean to know a certain word or phrase. It is a challenging task since, as it has mentioned before, there are several ways in which words can be known.

As Milton (2009) claims, there are, at least, two factors that should be considered when making vocabulary assessment materials: *reliability* and *validity*.

*Reliability* is generally defined as “the degree to which the result of a measurement, calculation, or specification can be depended on to be accurate” by the Oxford Dictionary.

So, regarding vocabulary assessment, a test would be reliable if it is stable over time. For instance, if a tester delivers the same test to the same person twice in a day, her or his knowledge cannot change significantly, then, the tests should show the same scores or grades in both samples. If so, the test could be claimed to be reliable.

On the other hand, *validity*, according to Oxford Dictionary, is “the quality of being logically or factually sound; soundness or cogency” or “the state of being legally or officially binding or acceptable.”

*Validity* addresses the issue of if a test really measures what it is supposed to measure, and it is a difficult field to be analysed.

Within this concept of validity we find *content validity*, which deals with this purpose, and considers whether a test has the necessary and appropriate items to measure vocabulary knowledge (Milton, 2009: p.18).

Tests of vocabulary breadth tend to use word lists from frequency information for testing. Nation’s (1990) Vocabulary Levels Test, for example, extracts words from the higher frequency bands in English from the University Word List (Nation, 1990). This type of assessment is thought to be generally content valid.

*Construct validity* considers whether the test measures the construct or skill it is supposed to, and it is closely related with the previous one. According to Milton (2009: p. 19),

For the measurements of productive knowledge, learners have to produce something, and if it is not a piece of writing or speech, what it can be? The requirement of researchers in the area of testing productive vocabulary knowledge is to elicit language that is truly representative of the learners’ productive vocabulary and which can be argued to have good construct validity (Milton, 2009: p. 19).

In order to assure the effectiveness of the content and construct validity, two different tests of the same quality are given to the same learner, and then, the results are compared. It is called *concurrent validity* test, and it supposed to help to validate a test’s construction and content.

The last point is *face validity*, which tries to find out if the test is “credible to users as a test of what it is supposed to measure” (Milton, 2009: p. 19)

Finally, we have to bear in mind the distinction between *objective* and *subjective* testing methods. The first one is considered to have high reliability after being test and retest.

Multiple-choice and forced answer exercises are some examples of these tests. On its part, subjective tests or assessment methods are less reliable and satisfactory because it uses written essays or open-ended comprehension questions, which involve not only vocabulary knowledge but also other skills such as mastering the grammar or choose the correct register in the foreign language.

#### 4.2 Word Frequency and Frequency Model

There are several factors that influence the learning of a word completely by a non-native speaker. For example, the sounds and the letters that form the word, the length, the similarities with his native language, how it can be inflected or derived, etc. and these factors, and others, can make a word difficult or easy to acquire (Milton, 2009: p. 22).

In addition, there is a more relevant factor which influences the learning process when dealing with word knowledge and it is *frequency*. The more a learner encounters a certain word in written texts or in oral speech, the better he or she will learn the word.

It is important for making vocabulary tests to know which words are likely to be learned and which words are not. This would make easier for researchers and testers to construct their assessment instruments.

Undoubtedly, in our daily speech people tend to use some words much more often than others, and even there are words that are almost never used by speakers. Hence, words such as articles *the* and *a/an*, prepositions such as *in* and *of*, conjunctions such as *and* and pronouns occur very frequently in English language. Therefore, learners are supposed to learn these kind of words first and get used to them more quickly (Milton, 2009: p.23).

Whereas function or grammatical words tend to appear millions of times in corpus, content or lexical words, which convey meaning, are much less frequent and, consequently, more difficult to acquire. Nevertheless, in order to master a language, both types of words (lexical and function words) are compulsory and necessary for learners to know.

As it has been stated above, there is a clear relationship between word frequency and the probability of a word to be learned by an individual because it tends to appear more frequently in any sort of text or oral speech. So it is an important concept to bear in mind when measuring second language vocabulary acquisition.

“This idea goes back 100 years ago to the pre-structuralist, scientific method in language teaching.” Palmer (1917: p. 123) wrote that “... the more frequently used words will be more easily learnt...”. Later, writers such as Mackey (1965) and McCarthy (1990) respectively accept this statement as an evidence (Milton, 2009: p. 25).

This assumption is widely accepted by researchers in this area, and, over the last years one of the benefits from this is the creation of a model that could be tested empirically. “Meara (1992) does this by graphing the relationship and producing a frequency profile” (Milton, 2009: p.25). This model has proved to be extremely stable.

Milton (2006a) conducted a study of all 227 learners at a language school in Greece, with abilities ranging from beginner to upper-intermediate learners of English, expressly to test the frequency hypothesis. The learners were given an orthographic vocabulary recognition Yes/No test with 20 test word from each of the first five 1000 words frequency bands which Meara includes in his model. The test used was X-Lex (Meara & Milton, 2003), [...]. The results, when graphed produced a profile, high on the left and tapering off to the right, as the Meara model suggested it should.

In other words, this study along with the ones conducted by Richards and Malvern (2007) or Aizawa (2006) among others, highlights the fact that the most frequent words in a language tend to be learned earlier than less frequent words.

Nonetheless, we have to be cautious about asserting this as a general rule, because young learners, for example, will learn the thematically important words in addition to the highly frequent words. So, words such as *giraffe* or *tiger* would be learnt in an early stage although these words are infrequent for adults.

Additionally, it is important to bear in mind that “frequency information does not provide information about difficulty. In fact, some of the most frequent words are the most difficult words for English learners” (Milton, 2009: p. 29).

Thus, this idea of frequency information and frequency lists has resulted in the development of frequency tests or frequency assessment tools to measure vocabulary knowledge in second language learners of English.

#### 4.3 Word Difficulty

As the evidences have shown, word frequency is one of the most important factors in understanding which words are more likely to be learnt, together with the idea of *word difficulty*. But, how can we classify a word as difficult or non-difficult?

Firstly, it is important to take into account the form of the word itself. That is to say, the spelling and the sounds involved phonologically. If the word contains difficult combination of sounds for the learner, it will be more difficult to learn and to recall to use it (Milton, 2009: p. 35).

Rodgers' (1969) study of English speaking students of Russian, for example, reported that words with non-English sound combinations and which were difficult to pronounce, were not learned as well as words which are easier to pronounce (2009: p.35).

Ryan (1997) points to writing confusions in English such as wells and wheels, left and lift and present and prison, where the consonants remain relatively unaffected by error, but the vowels are often miss-positioned, omitted or substituted (2009: p. 36).

Another factor to be considered is whether the word in the foreign language is similar to the word in the source language. Gairns and Redman (1986: p. 67) illustrates this idea with cognates such as *taxi*, *bar* and *hotel*. As these words are very similar in the majority of the European languages, they are easier to learn and to use in the target language, although the pronunciation may vary in some occasions.

Word length has also importance when talking about word difficulty. Shorter words are claimed to be learnt earlier and more quickly than longer words.

Gairns and Redman (1986) also suggest that “concrete items that can be represented visually, or demonstrated simply, may also be more economical to teach and learn than abstract items or ideas” (Milton, 2009: p.36). So words such as *head* or *car* would be easier to learn than more abstract concepts like *mind* or *superstition*.

To summarise, all the factors mentioned above may affect the process of learning and remembering a word in a foreign language. The length of the word, if the word is similar in both languages, concrete vs. abstract words and the form of the word itself would affect word knowledge.

## **5. Testing Vocabulary**

### **5.1 Testing and Assessment**



This section focuses on testing and assessment language level in general. First, there is a brief introduction on the topic, then the difference between formative and summative assessment is explained, in third place, there is a description of the different types of tests depending on the purpose, and finally, there is a short description of the stages of test development.

According to Hughes (2003: p. 4), many language teachers and researchers in general do not trust language tests at all. They claim that the process of teaching and learning would be better without testing. So, in some way, assessment and testing may interfere and come in conflict with the activity of teaching.

Nevertheless, tests are really helpful and necessary to provide objective information about learners' language ability. They are necessary, for example, when you want to go to an English-speaking country in order to work or to study, or when you want to join a language course. Tests are also of great importance to "provide information about the achievements of group of learners [...] to see how rational educational decisions can be made" (Hughes, 2003: p.4).

Testing, in spite of being the most common way, is not the only way of gathering language knowledge information. Here, we have to draw a distinction between *formative* and *summative* assessment.

*Formative* assessment are used by language teachers in order to analyse the progress of their students, and to know if they have learned what they are supposed to.

Informal testes or quizzes may have a part to play in formative assessment but also simple observation and the study of portfolios that students have made of their work. Students [...] carry out self-assessments in order to monitor their progress and modify their own learning objectives (Hughes, 2003: p. 5)

On the other hand, *summative* assessment "are used at the end of the term, semester or year to measure what students have been achieved both by groups and by individuals" (Hughes 2003: p. 5). In this case, formal test should be administered but they should not be assess in isolation.

Focusing now on types of tests, and following Hughes (2003), we have to differentiate between four different kinds of tests depending on the purpose: proficiency tests, achievement tests, diagnostic tests and placement tests.

### *Proficiency tests*

*Proficient* is defined as “well advanced in an art, occupation, or branch of knowledge” by Merriam-Webster dictionary. In the case of some proficiency tests, ‘proficient’ means “having sufficient command of the language for a particular purpose” (Hughes, 2003: p.11). This kind of test is produced to evaluate the learners’ skill in a certain language without any instruction beforehand. These tests measure what candidates can do or cannot do in a particular language. “Whatever the particular purpose to which the language is to be put, this will be reflected in the specification of test content” (Hughes, 2003: p.12)

Regardless of the content or the level of difficulty, all these proficiency tests have one common feature: “they are not based on courses that candidates may have previously taken” (Hughes, 2003: p. 12).

### *Achievement tests*

On the contrary, achievement tests are based on language courses and they are designed to measure how well a learner, or a group of learners, have been in reaching the goals proposed in a course. Within this, we can make a distinction between two types: final achievement tests and progress achievement tests.

*Final achievement tests*, are those that are delivered at the end of the course, term or semester and are intended to measure the final knowledge according to course content. These tests should only include the contents which appear in the course syllabus or on the books that have been used during the course. This is referred as syllabus-content approach and it is considered to be a fair test for students due to they know what may appear in the exam. The disadvantage is that, “if the syllabus is badly designed [...], the results of a test can be very misleading” (Hughes, 2003: p. 13). This problem could be overcome by basing the test content on the objectives of the course, this will offer more accurate and valid results.

The other type of achievement test is known as *progress achievement test*. As we can work out from his name, these tests are made for measuring learners’ knowledge through the course, in other words, assessing their progress. This kind of test should be related to objectives too. One way of putting these tests into practice is to deliver final achievement tests during the course, scoring the progress made by students. However, this approach may lead students to low marks, and therefore, they would feel depressed and discouraged.

One alternative could be to establish short-term objectives and then makes a clear progression towards a final achievement test.

### *Diagnostic tests*

These tests are intended to “identify learner’s strengths and weaknesses” (Hughes, 2003: p.15), and as a wide language skills it seem to be quite direct and unambiguous. The problem with this sort of test arises when teachers or testers want to evaluate a detailed analysis of a certain area of a language.

According to Hughes (2003), there is a very interesting web-based development called DIALANG which offers diagnostic tests in fourteen European languages and the test is divided into reading, writing, listening and grammar and vocabulary assessment. In that way, we could consider it as a very useful instrument for self-assessment and for being aware of our necessities when learning a new language.

### *Placement tests*

Placement tests are designed to supply knowledge information in order to place students into a course that fits better for their current skills or abilities in a language. “The placement tests that are most successful are those made for particular situations” (Hughes, 2003: p. 17). They can be constructed “in house” and administers can include any kind of items to measure the level of the participants.

Another interesting point when dealing with tests and language assessment is how tests are constructed. Regarding this issue and following Hughes (2003: p. 58), he recommends to follow the next steps when making a test:

1. Make a full and clear statement of the testing ‘problem’. What kind of test is to be? What abilities are to be tested?
2. Write complete specifications for the test. Information on content, test structure, timing, medium/channel, techniques to be used, criterial levels of performance and scoring procedures.
3. Write and moderate items. That implies sampling, writing items and moderating items.
4. Trial the items informally on native speakers (twenty or more, if possible) and reject or modify problematic ones as necessary.

5. Trial the test on a group of non-native speakers similar to those whom the test is intended under test conditions. This kind of trial is often not feasible.
6. Analyze the results of the trial and make any necessary changes. There are two kind of analysis: statistical and qualitative
7. Calibrate scales. This means, essentially, collecting samples of performance which cover the full range of the scale.
8. Validation.
9. Write handbooks for test takers, test users and staff.
10. Train any necessary staff (interviewers, raters, etc.)

(Hughes, 2003: p. 58)

## 5.2 Testing Vocabulary

This last section tries to offer a description of the main instruments for assessing vocabulary knowledge in English as a foreign language. In the first part, there is a brief distinction between recognition and production ability and then it provides some typical exercises to evaluate vocabulary together with a short information in each of them. Later on in the section, a number of questions and problems that may appear when dealing with these tests are explained. To finish, the section analyses the differences between vocabulary breadth and depth, giving some examples of each of them and providing an explanation that helps to distinguish these types of vocabulary knowledge.

Grammar teaching and assessment have had a great importance since a long time ago, and, for that reason, many tools and instruments to assess it have been developed throughout the ages, as well as teaching techniques and methodologies. It is widely acknowledged that teachers of languages in general have always tended to focus more on grammar mastery than on vocabulary acquisition.

On the contrary, in recent years, some authors have centred themselves in investigating vocabulary acquisition, teaching and assessment issues, because they thought it was a very important sphere in mastering a language. For instance, Hunt and Beglar (2005) argued that “the heart of language comprehension and use is the lexicon”. Likewise, Singleton (1999) claimed that “the major challenge of learning and using a language – whether as L1 or L2- lies not in the area of broad syntactic principles but in the ‘nitty-gritty’ of the lexicon.”

Besides, according to Read (2007), “the dominance of the communicative approach to language teaching in the past thirty years has thrown up various challenges to the validity of the conventional vocabulary test and this has prompted some re-thinking of the nature of lexical ability and how it can be assessed”.

Although measures of vocabulary knowledge have been included in reading assessment since the 1920s, the traditional means of assessing vocabulary have recently come under fire for being “driven by tradition, convenience, psychometric standards and a quest for economy of effort” (Pearson, Hiebert & Kamil, 2007: p. 282).

Vocabulary is thought to be easier to test than assessing other areas of language knowledge like grammar or discourse knowledge, due to the fact that the units that are measured can be separated with quite ease.

Following Hughes (2003: p. 180), he differentiates between testing *recognition ability* and testing *production ability*. The appendix 2 provides a set of examples of these types of test.

#### *Testing recognition ability*

It focuses on the capacity of learners of recognising a word in a foreign language. It is divided into three types of exercises. Recognising synonyms, definitions and choosing the correct word for a certain context. Some of these exercises are shown in appendix 1 as a way of illustration.

#### *Testing production ability*

Testing productive ability is much more difficult than recognition ability, consequently, it is practically never attempted in proficiency tests.

In the first exercise the candidate has some pictures and he or she has to write down their name. In the second one, there is a definition with a blank. This gap is the target word, and the learner is supposed to know it. And the last example is a sentence in which there is, at least, one gap to be filled with the correct word. Appendix 1 provides these three exercises on production ability as an example.

It is important to highlight the fact that none of the tests provide any options.

There are a wide range of different vocabulary item tests, and some of the most typically used based on Nation (2001: p. 345) are explained below. An example of each one is shown in the appendix 2.

*A 1000 word level true/false test (Nation, 1993)*

This test attempts to measure the most frequent 1000 words in English through a true/false questionnaire in which the participant has to write a T, a F or an X (if he/she does not understand the sentence).

*A vocabulary depth test (Read, 1995)*

As the title suggests, this test tries to measure vocabulary depth. In order to do this, there is a word and the participant has to choose one word from each of the two boxes that relates in some way with the target word. They have to choose, at least, one word from each box.

Another type of exercise gives the participant an incomplete definition of a certain word, and he or she has to fill the gap with a word chosen from a list on the right. There are some unnecessary words.

*A sensitive multiple-choice test (Joe, 1994)*

This test tries to measure vocabulary depth in a different way. It provides a word and five possible definitions. The participant has to choose the correct one.

*A translation test (Nurweni and Read, 1999)*

This kind of test is very simple. The participant is given a sentence in which a word is underlined. He or she has to translate this word into his or her first language.

*Yes/No tests*

Yes/No or checklist tests have been used and widespread since the last years of the 19<sup>th</sup> century, but, “it is enough to ask learners if they know the word?” (Nation, 2001: p. 364). Here is an example of a test extracted from Meara (1989).

**Tick the words you know.**

adviser _____	moisten _____
ghastly _____	patiful _____

contord \_\_\_\_\_

profess \_\_\_\_\_

implore \_\_\_\_\_

discard \_\_\_\_\_

The total score is calculated by removing the proportion of non-words that the candidate has marked as known from the number of real words marked as known. Meara and Jones (1990), Meara (1990b and 1991) and Meara and Buxton (1987) have used this sort of assessment with second language learners, and they state it to be reliable, valid and practical for the measurement of vocabulary knowledge in second language students (Nation, 2001: p. 348).

In a study conducted by Goodrich (1977), words of similar spelling were found not to function as distractors at all. Thus, it may be said that the employment of unreal words which are based on existing words in these tests are not distracting (Nation, 2001: p. 348).

The major problem of this sort of test is that learners or candidates do not show the real knowledge of a word that they possess.

Another recurring question that may arise is whether this kind of test should give options or they should not. Multiple-choice item tests have been administered widely due to the fact that they are easy to mark and less time-consuming than other tools for vocabulary assessment. They have also “a degree of respectability because they have been used in standardised tests like TOEFL” (Nation, 2001: p. 349).

It is generally assumed that item tests are the easiest to answer for first language learners (Nist and Olejnik, 1995; Paul, Stallman and O’Rourke, 1990). In order to overcome this issue, Nagy, Herman and Anderson (1985) claimed that it was possible to construct multiple-choice item test with different degree of difficulty by changing the closeness in meaning between non-words and real words. Obviously, this kind of test inspires candidates to guess the correct answer.

An interesting research was conducted by Paul, Stallman and O’Rourke (1990) regarding the strategies first language learners use to answer a particular multiple choice test items, and then, they separated them into the next sections.

- Knowing the answer: the answer was chosen because learners said they knew it was correct.
- Association: the answer was chosen because it could be related in some way to something they knew about the word.

- Elimination: the answer was chosen by ruling out the other choices.
- Position of the options: the answer was chosen because it was first, last or in the middle.
- Readability of the options: the answer was chosen because it was the only one they could read and understand
- Guessing: the learners did not know why they chose an answer or they said they just guessed.

(Nation, 2001: p. 349)

The results of the research demonstrated that guessing “is not a major problem with multiple-choice items and that learners’ responses are generally not random but largely driven by some knowledge of the words”, for example association (Nation, 2001: p. 350).

One of the advantages of this test is that it can focus on particular meaning. However, the major drawback is that the words measured within the same block can affect each other. Additionally, Campion and Elley (1971) found that “changing the block a word was placed in often resulted in a big change in the number of correct answers” (Nation, 2003: p. 350).

Another subject that can be matter of debate is the possibility of using translation in vocabulary tests. According to Nation (2001), first language translation has always been considered as an incorrect method for testing and learning a foreign language. As far as I am concerned, language teachers try not to use translation as a way of testing their students, although, in my view, it is a very accurate and efficient way to have access to your students’ real knowledge of a word.

Besides, we have to bear in mind that, “the use of the first language meaning is like choosing a synonym, whereas a second language definition often involves a definition form [...] which requires greater grammatical skill” (Nation, 2001: p. 351). This type of assessment (first language translation) provides not only receptive and productive knowledge, but also recall and recognition of items.

In this last part of the work, there is a distinction between vocabulary breadth and vocabulary depth and how it can be measured. In my view, it is one of the main points when dealing with vocabulary knowledge and assessment.

### *Testing Vocabulary Breadth*



Research in the field of vocabulary size, also referred as vocabulary breadth, has had a great development since a long time ago. This type of measurement is used both for native and non-native speakers as well.

Some of the most typical examples of this kind of tests are checklist models, where participants have to recognise certain set of words in order to estimate his or her vocabulary knowledge mainly through passive recognition

Additionally, vocabulary size is closely related with reading comprehension skills because it can reveal the amount of difficulty a learner could have when facing a real example of some reading materials.

These tests usually need a large amount of sample of words taken from word frequency lists, and the answers have to be a simple task to show only if the participant knows the word or not.

Examples of frequency lists in where researchers can base their tests are, among others: the British National Corpus (Leech, Rayson and Wilson, 2001), which is the most accessible list available in paper formant and online, Nation's, the aging General Service List (West, 1953), and the Academic Word List (Coxhead, 2000) which complements and improves the GLS.

However, in spite of the fact that there are many word lists available in books, papers and online, there is not a 'perfect' word frequency list because each one has their advantages and downsides. So, we can state that further investigation is needed to reach a unique, common and reliable frequency list from which tests can be made of.

Once a certain frequency list has been chosen, a vocabulary size test has to be designed from selecting a sample of target words of the mentioned list. One of the most broadly used to measure vocabulary size for second language learners is Nation's Vocabulary Levels Test (Nation 2001: p. 416-424). It asks participants to match words with their synonyms or definitions. "Nation's new Vocabulary Size Test (Nation & Gu, 2007) has a multiple-choice format, which each target word presented in a short non-defining sentence followed by four possible definitions" (Read, 2007: p. 110). Both types show the actual knowledge of each word from direct evidence.

These type of tests were first referred as checklist, but now it is widely acknowledged as Yes/No assessment tests. Anderson and Freebody (1983) included an important change

to this type of test; they introduced some unreal words in order to adjust the scores obtained by test-takers, assuming that, in some cases, guessing plays an important role when marking a word as known. This was a very satisfactory tool and more accurate and valid scores results from the use of this calculation.

Meara and some of his colleges, first in London and then in Swansea University (Meara, 1992; Meara & Milton, 2005) have developed Yes/No test and have made them available for practical use. This sort of tests are easy to computerize and administer, so it is seen as a great advance in assessment field (Read, 2007: p. 111).

One programme, X\_Lex, covers the first 5000 most frequent words of not only English, but also French, Spanish, Swedish and Portuguese, whereas Y\_Lex samples vocabulary in the 6000 – 10000 word range but just for English (Read, 2007: p. 111). [...] A further development is a version of X\_Lex in which the words are presented orally, this test is known as Aural\_Lex (Milton & Hopkins, 2005).

It seems necessary to research deeply in this area because as Milton and Riordan (2007: p. 132) claim, whereas the printed form of a word is relatively fixed, the spoken form can vary according to factors such as the linguistic context, the accent, and the possibility of misleading words with similar sounds (Read, 2007: p. 111).

Here, there is a reduced example of X\_Lex Vocabulary test.

Each one contains 20 items selected randomly from each of the first five 1000 words frequency bands in English. It also contains 20 false words [...]. Words from the first 1000 word band are presented in column 1. Words from the second 1000 word band are presented in column 2, and so on. False words are presented in column 6.

That	Both	Cliff	Sandy	Lessen	Darrock
With	Century	Stream	Military	Oak	Waygood
Before	Cup	Normal	Impress	Antique	Kennard
Person	Discuss	Everywhere	Staircase	Chart	Gazard
Feel	Park	Deny	Daily	Limp	Fishlock
Round	Path	Shot	Essential	Permission	Catileen
Early	Tower	Refer	Associate	Headlong	Gillen
Table	Weather	Independent	Conduct	Violent	Pardoe
Question	Wheel	Feeling	Relative	Fade	Frequid
Effect	Whole	Bullet	Upward	Rake	Hobrow

Market	Perform	Juice	Publish	Trunk	Candlin
Woman	Pity	Nod	Insult	Mercy	Litholect
Stand	Probable	Gentle	Cardboard	Anxious	Gumm
Believe	Signal	Slip	Humble	Pedestrian	Alden
Fine	Dish	Diamond	Contract	Arrow	Teadaway
Instead	Earn	Press	Mount	Feeble	Sumption
Produce	Sweat	Provide	Tube	Sorrow	Horozone
Group	Trick	Drum	Moreover	Brighten	Hyslop
Arrive	Manage	Reasonable	Crisis	Dam	Manomize
Difficult	Mud	Boil	Jug	Outlet	Horobin

Table 2. X\_Lex Vocabulary Test (Milton, 2009: p. 254)

Scoring is as follows: award 50 for each real word checked by the testee and total these to produce a 'raw score'. From the raw score deduct 250 for each false word, which is checked to produce an adjusted score and the estimate of words known (Milton, 2009: p. 254).

### *Testing Vocabulary Depth*

It is widely acknowledged that measuring vocabulary depth is a much more complex task than measuring vocabulary size or breadth. Thus, there is less progress in this area of vocabulary assessment and a few tools exist in order to assess depth of knowledge.

As I have mentioned in previous sections, there are various authors who have analysed the components which imply knowing a word and have attempted to know how it can be assessed. Depth generally includes "the shades of meaning a word may carry, its connotations and collocations, the phrases and patterns of use it is likely to be found in, and the associations the word creates in the mind of the user" (Milton, 2009: p. 148).

Studies such as the one of Meara (1982) suggest that the word associations produced by second language learners are qualitatively different from those produced by native speakers.

In addition, Meara and Wolter (2004: p. 95) say, "we might find learners with different with similar vocabulary sizes, but very different degrees of organization in their lexicon". Consequently, there are learners who have a lot of words storage in their minds but poor organization, and the other way round. This may be the reason why learners with the same amount of vocabulary knowledge perform differently in the same academic examination (Milton, 2009: p. 150).

In another study Read (2009: p. 150) suggests that “vocabulary depth may not really be a single dimension. It is hard to see what principle unifies collocational, associational knowledge, constraints on use, polysemy and the other qualities that are placed within this dimension” (Milton, 2009: p. 150).

Now I will try to revise some of the tests that are available to measure vocabulary knowledge in terms of depth. In order to do this I will follow Milton (2009: p. 151-168).

One approach is the Vocabulary Knowledge Scale (Paribakht & Wesche, 1996) which is a word knowledge test based on asking learners not only if they know or recognise the word but also how well they know these words. The VKS presents participants a list of target words and a 5-point scale from ‘I don’t remember having seen this word before’ to ‘I can use this word in a sentence’. If they think they know a certain word then they have to use the word in context or translate it.

<ol style="list-style-type: none"><li>1. I don’t remember having seen this word before.</li><li>2. I have seen this word before but I don’t know what it means.</li><li>3. I have seen this word before and I think it means_____ (synonym or translation).</li><li>4. I know this word. It means_____ (synonym or translation).</li><li>5. I can use this word in a sentence: _____ (If you do this section please complete number 4)</li></ol>
--

Figure 1. Wesche and Paribakht’s VKS (1996)

Another attempt of measuring vocabulary depth is to choose a single element of this quality, for example an idiom or a collocation and assesses it in isolation. There are two actual instances of this quality tests developed by McGavigan (2009) and Gyllstad (2007) respectively.

On the one hand, McGavigan focus on measuring learners’ knowledge of fixed idioms. He uses frequency data on English idioms available in the Cobuild Dictionary of Idioms (Collins Cobuild, 1995), taking from them sample of idioms for testing purposes. Twenty idioms randomly chosen from each of the four frequency bands and gap-fill format questions are created for each idiomatic expression. Here there is an example of the test.

## Idioms Test

This is a test of Idioms Knowledge in native speakers of English. For the purposes of this test and idiom is FIXED phrase which is used metaphorically to describe a situation or feeling.

## Instructions

Please complete the following test items providing ONE word from each gap. Write answer in the box [...].

## Example

	Question	Answer
1	Look at the weather. It is raining cats and .....!	dogs

The results that emerge from the test are quite reliable according to McGavigan, besides, it may seem that foreign language learners tend to know much less idiomatic expressions than any native speaker even if learners have been living or working in an English-speaking country or if they have an advanced level of the language.

Two observations are worth considering. The first one is that knowledge of idioms correlate significantly with a measure of vocabulary breadth. The results suggest that it is necessary a minimum of around 3000 words before idioms knowledge is able to develop. And the second one is that it appears to be a frequency effect in learning idioms (Milton, 2009: p. 154).

On the other hand, Gyllstad focuses on word collocations. He tries to develop a test that can give reliable and valid data about this area of vocabulary knowledge and how it progresses.

Gyllstad (2007) produces two tests, *Collex* and *Collmatch*.

Collex presents learners with a larger number (50) of collocations. Alongside these real collocations are 50 pseudocollocations; combinations of words which are not collocations. The learner has to select the acceptable collocation.

In the Collmatch format, a series of grids are presented that invite the learner to match three verbs with six noun phrase objects. Learners are asked to tick the combinations they believe they can use in English (Milton, 2009: p. 155-156).

The items used to construct these tests are taken from the most frequent 3000 word range of the British National Corpus. As in the previous test, it appears compulsory to have a great amount of vocabulary size and a good basic competence in order to answer these kind of test successfully. It arises from the scores that participants are better at marking real collocations than at discarding false ones. The results of both tests (Collex and Collmatch) correlate very well and they “appear to function well within the limitations of the qualities they set out to measure” (Milton 2009: p. 157)

Tell a prayer	Say a prayer	<input type="checkbox"/>
Pay a visit	Do a visit	<input type="checkbox"/>
Run a diary	Keep a diary	<input type="checkbox"/>
Do a mistake	Make a mistake	<input type="checkbox"/>

Figure 2. Examples of Collex format (Gyllstad, 2007)

	<i>Charges</i>	<i>Patience</i>	<i>Weight</i>	<i>Hints</i>	<i>Anchor</i>	<i>Blood</i>
Drop						
Lose						
Shed						

Table 3. Examples of Collmatch format (Gyllstad, 2007)

A simple conclusion might derive from these tests: vocabulary depth is very closely related with vocabulary breadth, because a sizeable vocabulary is necessary to build a large number of connections between words.

This kind of test, alongside the others mentioned before, has some advantages and disadvantages. For example, it is very difficult to generalise about the progress of learners’ vocabulary depth since test items are not standardised. On the contrary, different studies of correlation conducted by Wesche and Paribakht demonstrate that this method is very reliable and valid.

Apart from that, there is another type of tests called *association tests*, which tries to combine tests of different qualities to get a better result of vocabulary depth at once.

The first attempt was to create a way of measurement through an interview. In the figure below and example of this kind of test is shown.

<p>TO INTERPRET</p> <p>1. Write two sentences A and B. In each sentences use the words given.</p> <p>A.     interpret                      experiment</p> <p>B.     interpret                      language</p> <p>2. Write three words that can fit in the blank.</p> <p>to interpret a (n) _____.</p> <p>   _____.</p> <p>   _____.</p> <p>3. Write the correct ending for the word in each of the following sentences.</p> <p>Someone who interprets is an interpret_____.</p> <p>Something that can be interpreted is interpret_____.</p> <p>Someone who interprets gives an interpret_____.</p>
--

Figure 3. Test sheet for the word Interpret (Read, 2001: p. 179)

This is a useful method of eliciting word knowledge because the learners are asked to know different shades of word meaning through different tasks. Additionally, the test evaluates whether the participant knows or does not know several derivational words from the word given. All the items used to construct these tests are taken from the University Word List. The problem of this test is that they need a lot of time to make them and a little number of words can be examined in this way (Milton 2009: p. 162).

Read (1995: p. 14) concludes that the test “can be seen as very efficient testing instrument and one that has the potential to focus on several different aspects of the meaning of the target word”. Wolter (2005: p. 37) was not so optimistic and suggested that “it is still not possible to suggest with confidence that this test has succeeded in [...] assessing depth of word knowledge”.

Another approach is the one proposed by Wolter and Meara (2005) and their V\_Links. They “try to count the number of links of all kinds rather than worry about what the links are” (Milton 2009: p. 163). By doing so, they expect to gather the whole depth dimension in a single assessment.

Undoubtedly, there are several words in English that are highly connected. For example a verb such as *get* or *take*, will match with a lot of different prepositions to make phrasal verbs. And there are also much more restricted words such as *kith* or *dint* (Milton, 2009: p.163).

Wolter's (2005) V\_Links test is attempt to operationalise this idea. In his test, 10 words are randomly selected from JACET8000 wordlist (Ishikawa *et al.*, 2003), and these words are presented on a computer screen in a circle so testees can use the on-screen pointer to visually link any word in the set to any other word.

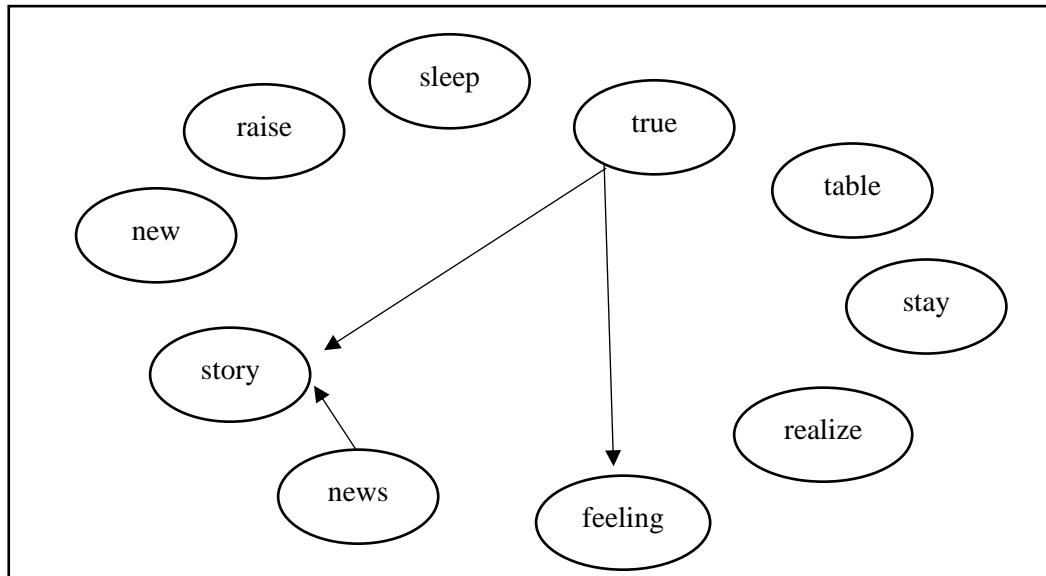


Figure 4. V\_Links version 3.0 (Wolter, 2005: p. 144)

Although it may seem as a very feasible and valid method for testing vocabulary depth, there are a number of problems too. One drawback is that there is no option of checking the quality of the associations involved. So, once the learner realises that the idea of the test is to link as many words as possible, the test stops working.

Henriksen (2008) encapsulates the problem in her diagram of the three types of vocabulary knowledge to be found in semantic memory. It looks like this:

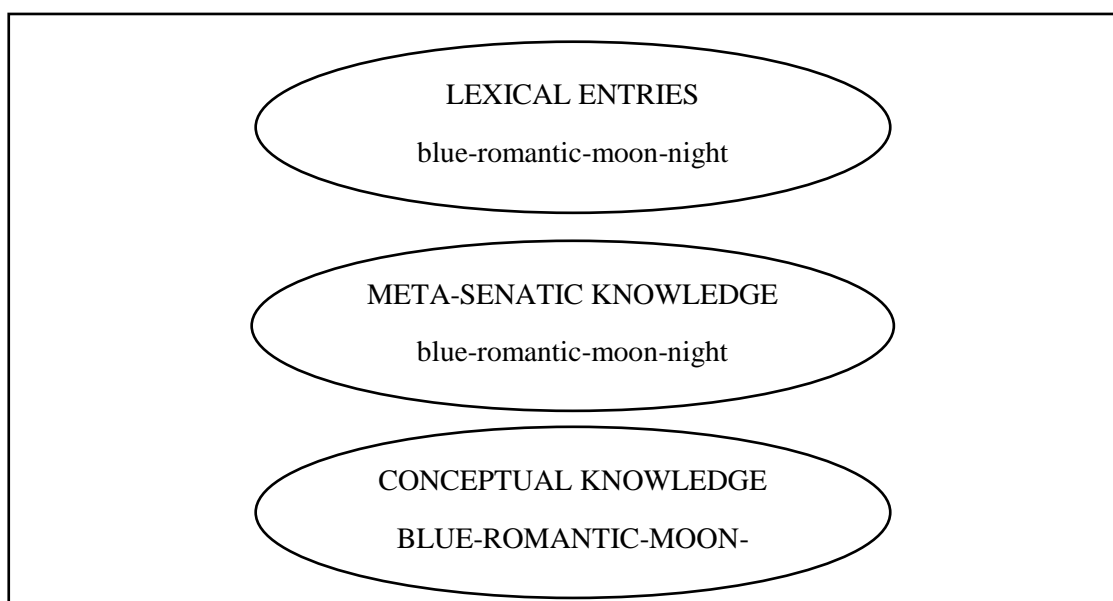




Figure 5. Links between and within the three levels of lexical representation (Henriksen, 2008: p. 29)

Level I contains conceptual or encyclopaedic knowledge. Level II contains a mental inventory of lexical items or words in whatever form, phonemic or graphic. Level III comprises meta-linguistic knowledge of the semantic relations between lexical items (Milton, 2009: p. 167).

To conclude, we can state that there is not a unique method to assess depth knowledge and that none of the tests mentioned in this section are workable and valid completely. Measurement of idioms and collocational knowledge seem to be quite useful and accurate for this purpose regarding the results and evidences. However, the other methods are not so firmly established and show some drawbacks although they can be used as an important source of information of vocabulary knowledge.

According to Milton (2009: p. 169), “whatever the qualities of depth and breadth are, they are linked, and qualities of depth really seem to appear only after a sizeable vocabulary breadth has been attained”.

## **6. Conclusion**

The aim of this work has been to describe and explain the main instruments or tools that are available for measuring vocabulary knowledge. But, in order to do so, this work provides the necessary background to a better understanding of further concepts and ideas. Although the main point is instruments for assessing vocabulary knowledge, there is also a brief commentary on general tests and assessment of language. This work would be useful for teachers of languages and researchers in the field of language acquisition and measurement. It has attempted to describe the different instruments available depending on the area of vocabulary that we want to measure exactly. Further research is needed in this area of knowledge in order to find a ‘perfect’ test that measures all the spheres of vocabulary at once. In addition, teachers should use this kind of tools consciously, basing on the area they want to measure and scoring the results accurately and they should realize that there are many ways to assess their students. The limitations are, basically, time. Making better tests is possible but it is very time-consuming for both construction and scoring.

## 7. References

Boogards, P. & Laufer, B. (2004). *Vocabulary in a Second Language*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cambridge University Press. (2008). *Cambridge Online Dictionary*.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishing.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

*Introductory Guide to the Common European Framework of Reference for English Teachers*. 2013, de Cambridge University Press. Sitio web: <http://www.englishprofile.org/images/pdf/GuideToCEFR.pdf>

Jiménez, R. M. (2002). El Concepto de Competencia Léxica en los Estudios de Aprendizaje y Enseñanza de Segundas Lenguas. *Universidad de La Rioja*, pp. 150-162.

Meara, P. (1996). The Dimensions of Lexical Competence. *Performance and competence in second language acquisition*. Cambridge: Cambridge University Press. 1996. 3553.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

Miralpeix, I. & Meara, P. (2016). *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Read, J. (2007). Second Language Vocabulary Assessment: Current Practices and New Directions. *International Journal of English Studies*, pp. 105-125.

Sánchez, A. & Manchón, R. (2007). Research on Second Language Vocabulary Acquisition and Learning. *International Journal of English Studies*, pp. vii-xvi.

Stuart, C. P. (1999). *An Introduction to Linguistics*. New York: St. Martin's Press.

Tanaka, S. (2008). Lexical Competence in English. *COCONE Institute for Language Education*.

Trask, R.L. (1999). Key Concepts in Language and Linguistics. London: Routledge.

## Appendix 1

### *Recognise synonyms*

**Choose the alternative (a, b, c or d) which is closest in meaning to the word on the left page.**

Gleam: a) gather b) shine c) welcome d) clean

Shine: a) malm b) gleam c) loam d) snarl

The problem of this test is what distractors should be used.

### *Recognise definitions*

loathe means: a) dislike intensely b) become seriously ill  
c) search carefully d) look very angry

Bill is someone I loathe: a) like very much b) dislike intensely  
c) respect d) fear

### *Recognise appropriate word for context*

The strong wind \_\_\_\_\_ the man's efforts to put up the tent.

a) disabled b) hampered c) deranged d) regaled

### *Pictures*

**Each of the objects drawn below has a letter against it. Write down the names of the objects.**

a) \_\_\_\_\_ c) \_\_\_\_\_

b) \_\_\_\_\_ d) \_\_\_\_\_

### *Definitions*

A \_\_\_\_\_ is a person who looks after our teeth.

\_\_\_\_\_ is frozen water.

\_\_\_\_\_ is the second month of the year.

*Gap filling*

One or more sentences with a single word missing.

Because of the snow, the football match was \_\_\_\_\_ until the following week.

I \_\_\_\_\_ to have to tell you this, Mrs Jones, but your husband has had an accident.

## Appendix 2

*A 1000 word level true/false test (Nation, 1993)*

**Write T if a sentence is true. Write F if it is not true. Write X if you do not understand the sentence.**

1. We cut time into minutes, hours and days. \_\_\_\_
2. Some children call their mother Mama. \_\_\_\_
3. All the world is under water. \_\_\_\_
4. When you keep asking, you ask once. \_\_\_\_

*A vocabulary depth test (Read, 1995)*

**Choose four words that go with the test word. Choose at least one from each of the two boxes.**

Sudden:

Beautiful	Change
Surprising	Noise
Quick	Doctor
Thirsty	School

**Choose one word from the list on the right to complete the sentence. Do not use the same word twice.**

1. A journey straight to a place is \_\_\_\_\_. faint
2. An illness that is very serious is \_\_\_\_\_. acute
3. A river that is very wide is \_\_\_\_\_. common
4. Part of your body that is not covered by any clothes is \_\_\_\_\_. bare

5. Something that happens often is \_\_\_\_\_.

alien

broad

direct

*A sensitive multiple-choice test (Joe, 1994)*

**Circle the choice that best gives the meaning of the underlined word.**

chronic means: a) lasting for a long time

b) dissatisfied

c) to greatly decrease

d) effective and harmless

e) don't know

*A translation test (Nurweni and Read, 1999)*

**Translate the underlined words into your first language.**

1. You can see how the town has developed.

2. I cannot say much about his character.

3. Her idea is a very good one.

4. I want to hear only the facts.